The CDF Data Acquisition System for Run II

Arnd Meyer
Fermilab

Computing in High Energy Physics 2001
Beijing, September 2001

Outline

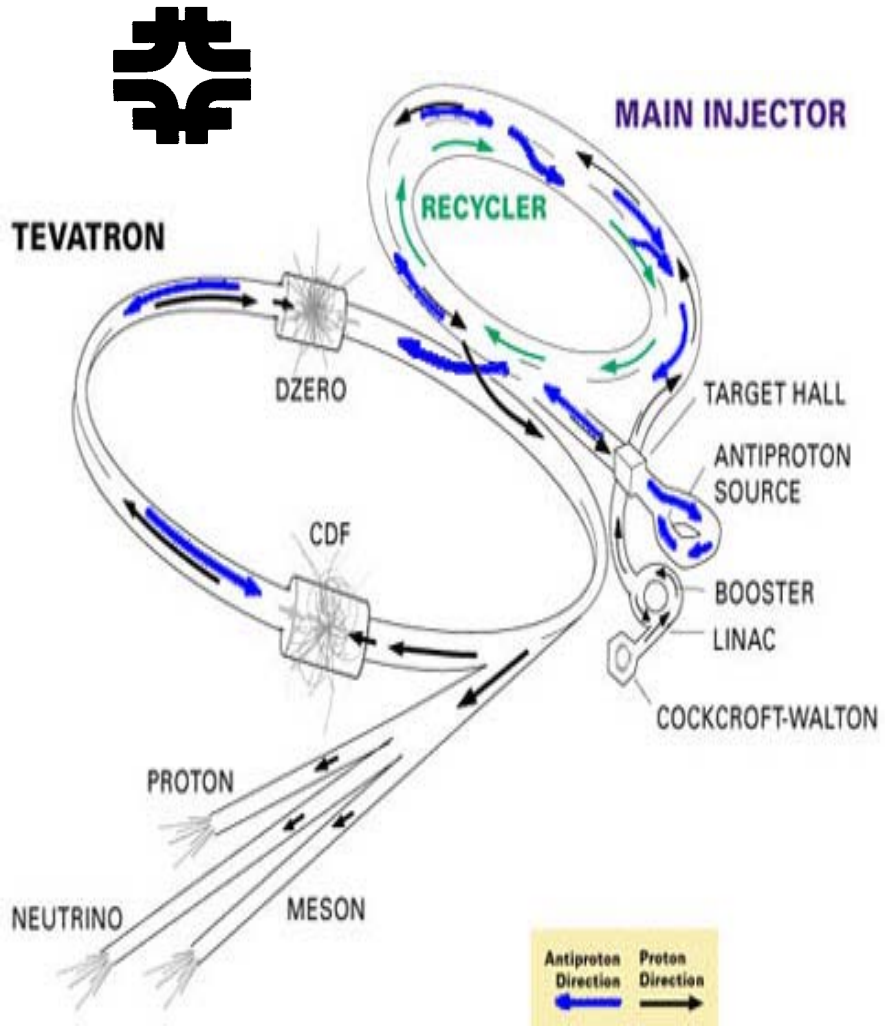
Arnd Meyer
Sep 5, 2001

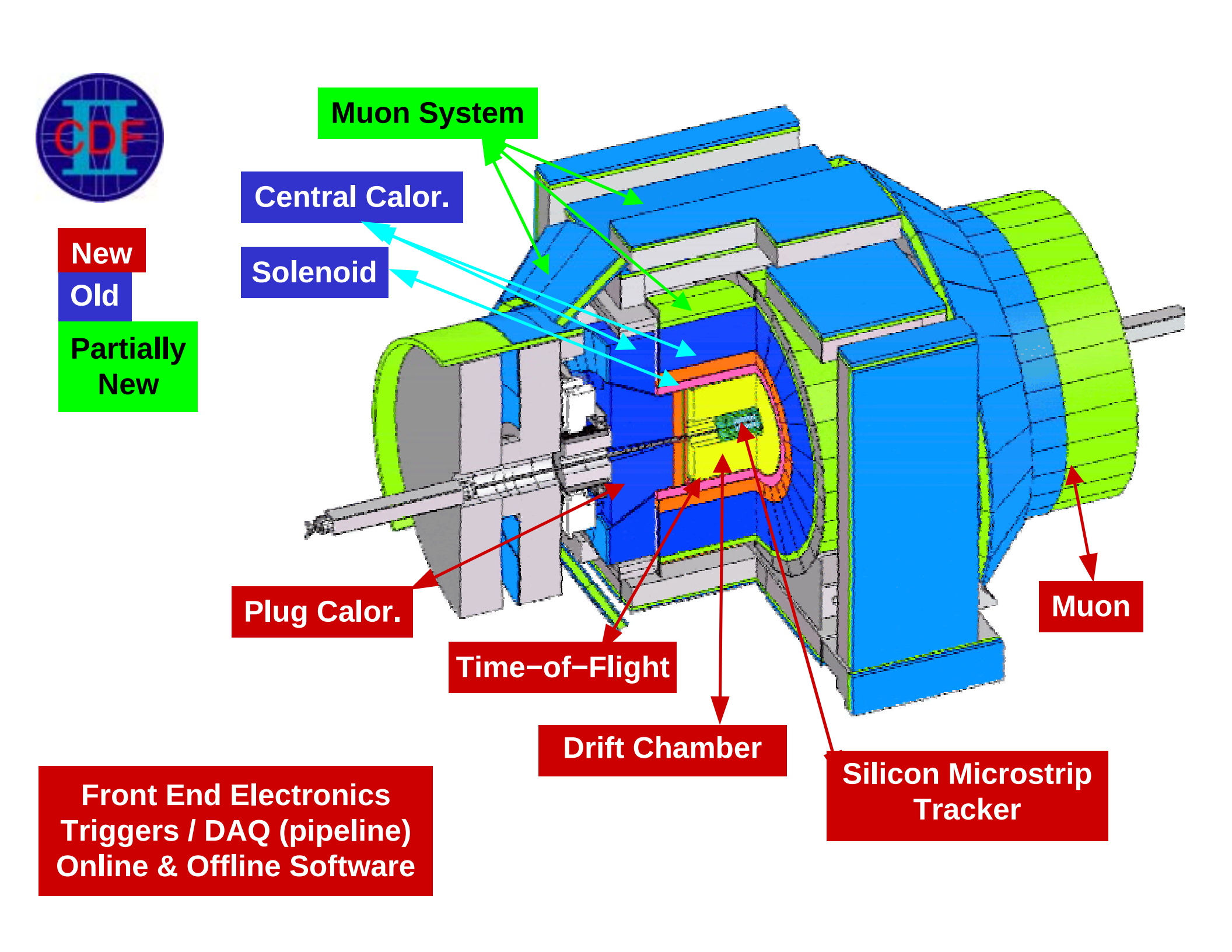
- Introduction
- Architecture and general features
- Front-End crates
- Trigger
  - Levels 1 and 2, example track triggers
  - Event Builder
  - Level 3 farm
- Silicon DAQ
- Consumer Server / Data Logger
- Run Control and related online software
- Commissioning and Performance
- Summary

The Fermilab Accelerator Complex

Arnd Meyer
Sep 5, 2001

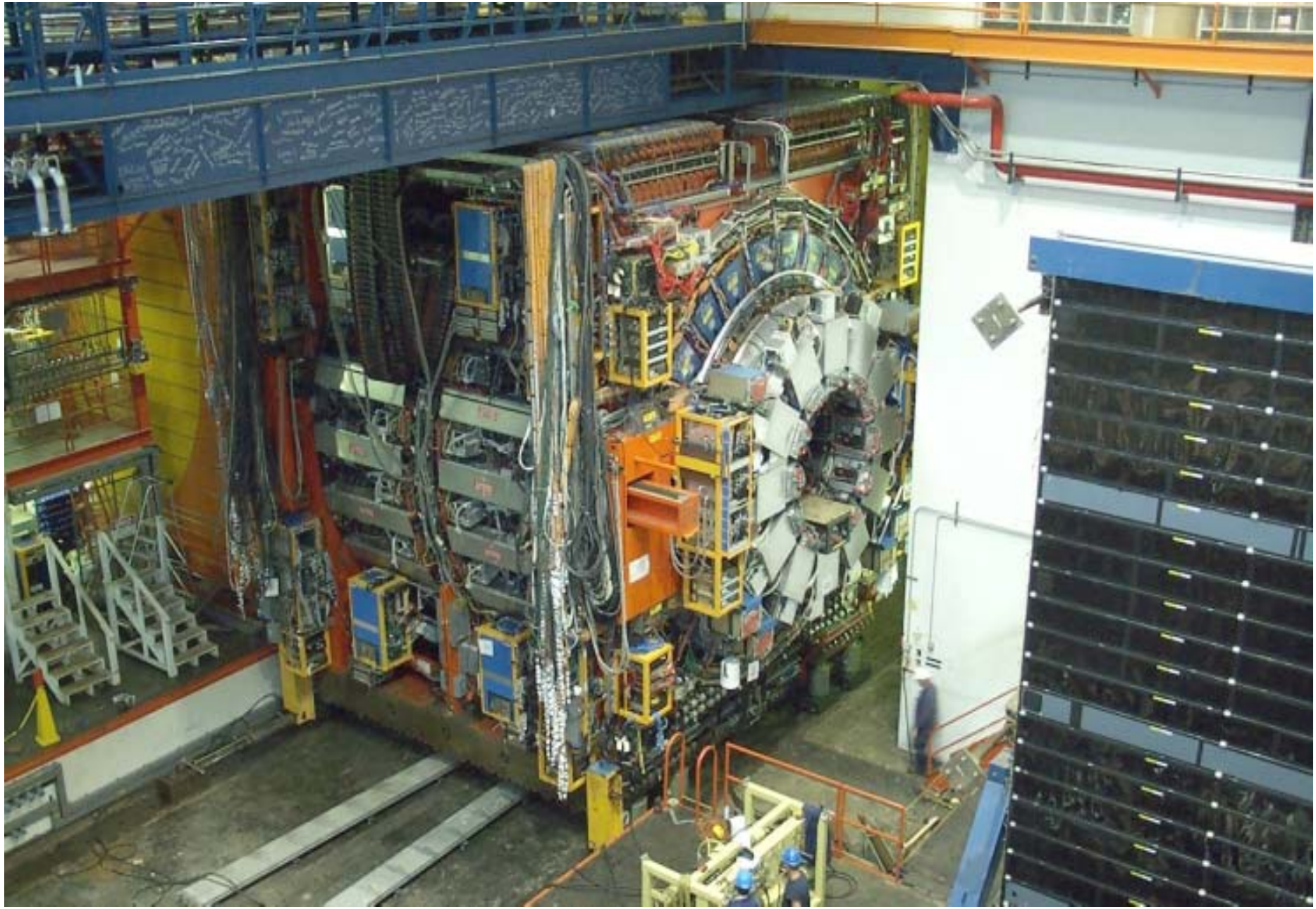
- Run IIa (2001 – 4): 2 fb^{-1}
  - Main Injector : x5
  - 150 GeV proton storage ring replaces Main Ring, the original Fermilab accelerator.
  - Recycler : x 2—3 (2003 – 4)
  - Re-cools p-bar from Tevatron
- Run IIb (2005 – 7): 15 fb^{-1}
  - electron cooling, crossing angle, electron lens: x 2—3
- Increased # of p and p-bar bunches:
  - 6 (3500 ns) → 36 (396 ns) → ~100 (132 ns)
- Higher energy collisions:
  - E_{proton} = 900 → 980 GeV
- Typically factor 200 in statistics from accelerator upgrades!
- Plus detector upgrades / improved acceptance ...




Muon System
Central Calor.
Solenoid
New
Old
Partially New
Plug Calor.
Time−of−Flight
Drift Chamber
Silicon Microstrip Tracker
Muon
Front End Electronics
Triggers / DAQ (pipeline)
Online & Offline Software

Detector Roll–In

Arnd Meyer
Sep 5, 2001



Key Requirements

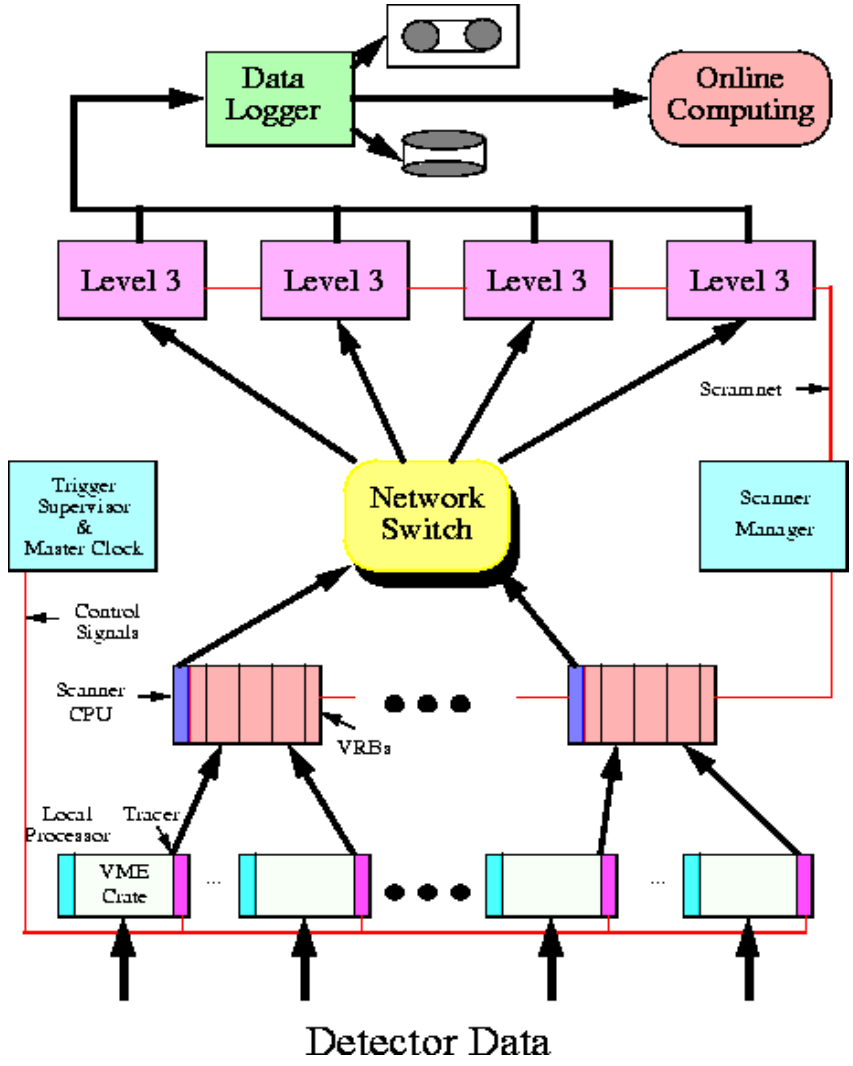
Arnd Meyer
Sep 5, 2001

- $\mathcal{L} \sim 2 \times 10^{32}$ cm^{-2}sec^{-1} (20 x run I)
- Read out ~1M readout channels (10 x run I) at 300Hz-1kHz
- Event size ~250kB
- Permanent logging at 75Hz, 20MB/sec to tape/mass storage
- Partitionable
- Control and monitoring at all levels
- >90% live
- Bunch crossing every 132ns (currently every 396ns, 1.7MHz)
- Tevatron stores last up to 50h, little or no quiet time between stores

System Components

Arnd Meyer
Sep 5, 2001

- Trigger Supervisor and Crosspoints
  - Interface trigger system with DAQ
- Front-end and trigger VME crates
  - Most electronics, lowest level readout
- Event builder
  - Assemble event fragments
- Level3 trigger
  - Format event, final trigger decision
- Consumer Server/Logger
  - Write data to disk, distribute to online monitoring programs
- DAQ control and monitoring programs
- Event data monitoring programs ("Consumers")



Front−End Crates

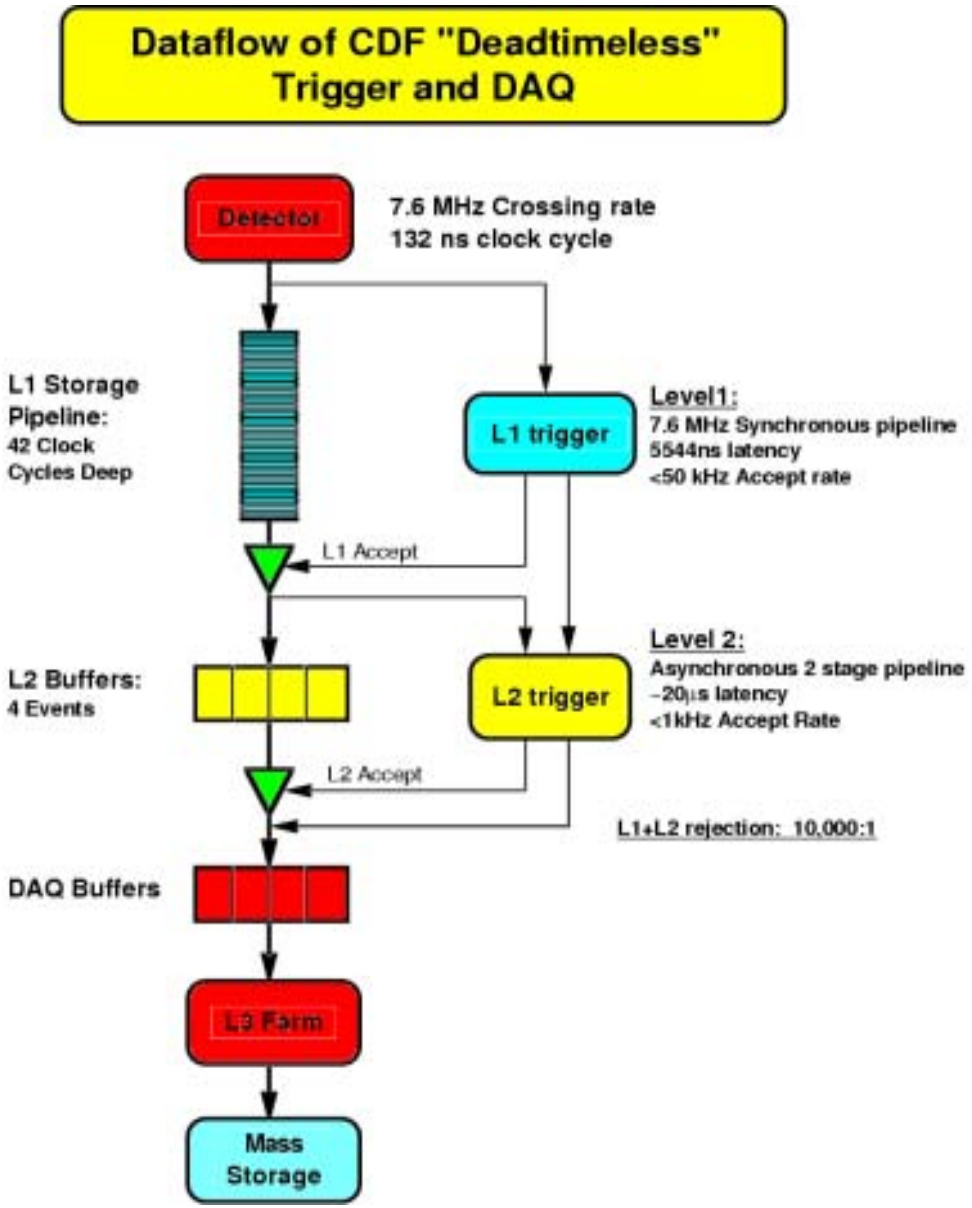
Arnd Meyer
Sep 5, 2001

- Front-end and trigger electronics are housed in ~125 VIPA VME crates, 21 slots 9U x 400mm
  - ~half on detector, ~half in counting rooms
- Over 1700 main modules of about 60 types (+ >400 spares)
- Each module has standardized registers/memory blocks for event data readout, configuration parameters
- Over 1000 transition (I/O) modules of about 25 types
- 60 - 6U Eurocard crates with >700 modules for Showermax readout and clock system
- Over 25000 daughter boards
- Each crate has PPC based crate controller (Motorola MVME2301 or better)
  - Runs VxWorks real-time OS
  - CDF written software to configure local VME modules, read out event data, and provide status information
  - Communication to run control and monitoring through "SmartSockets"
- Most crates also have a "Tracer" module
  - Receives signals from Trigger Supervisor and fans out on backplane for modules to pick up
  - Receives bunch crossing clock signals from MasterClock and fans out on backplane
  - Sends current readout status lines back (DONE, ERROR, BUSY)
  - Optical data link to event builder
  - Silicon system different

Trigger Overview

Arnd Meyer
Sep 5, 2001

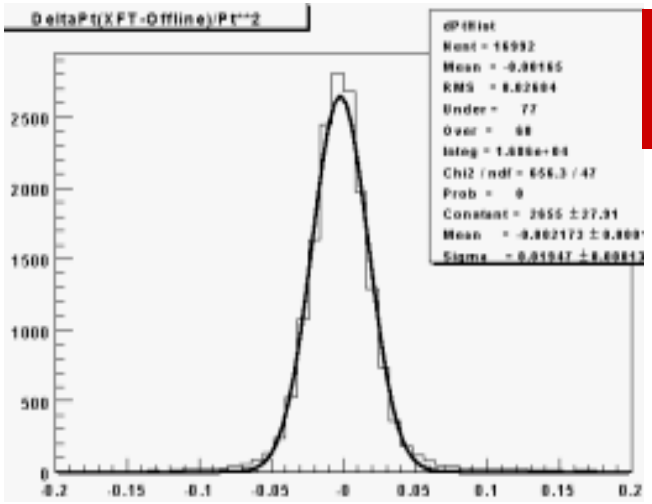
- Level 1:
  - Every front-end system stores data for 42 crossings
  - "Hardware trigger"
  - 50kHz accept rate
  - On L1 accept, data is stored in one of four L2 buffers
- Level 2 (asynchronous):
  - 20μs decision time
  - "Mostly hardware" trigger
  - Trigger algorithms run on custom Alpha boards
  - Displaced vertex trigger, improved matching, calorimeter clusters, ...
  - 300Hz accept rate (→ 1kHz)
- Event readout starts on L2A
- "Deadtimeless"
  - Deadtime only incurred when all L2 or DAQ buffers are full



L1: eXtremely Fast Tracker

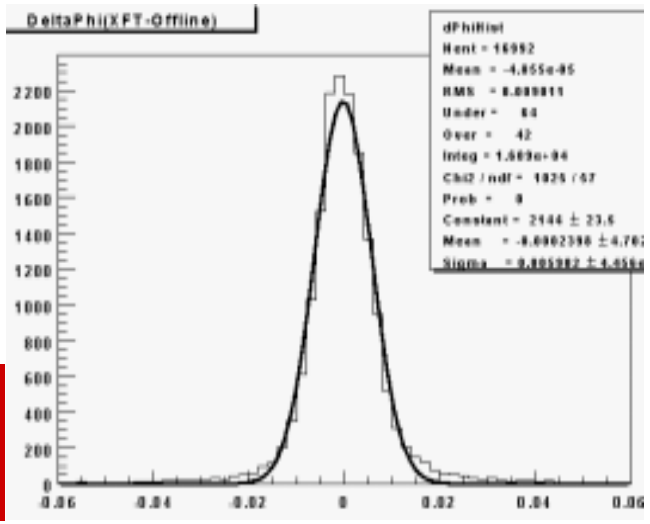
Arnd Meyer
Sep 5, 2001

- Level 1 consists of custom hardware, fully pipelined, uses information from central drift chamber, calorimeters, muon systems, luminosity detectors
- Multi-object triggers and matching between tracks and calorimeter objects / muon stubs
- High efficienty/purity track trigger on Level 1
- XFT receives prompt/delayed hits
- Finder modules identify track stubs in axial superlayers
- Linker modules "links" patterns of pixels to tracks
- XTRP system sends tracks to L1 muon, L1 calorimetry, and L1 track trigger, and on L1A to L2 systems
- Resolution comparable to offline tracks



$\Delta pt/pt^2 = 0.016\ GeV^{-1}$
Goal : 0.02 GeV^{-1}

$\Delta\phi = 6$ mrad
Goal : 8 mrad

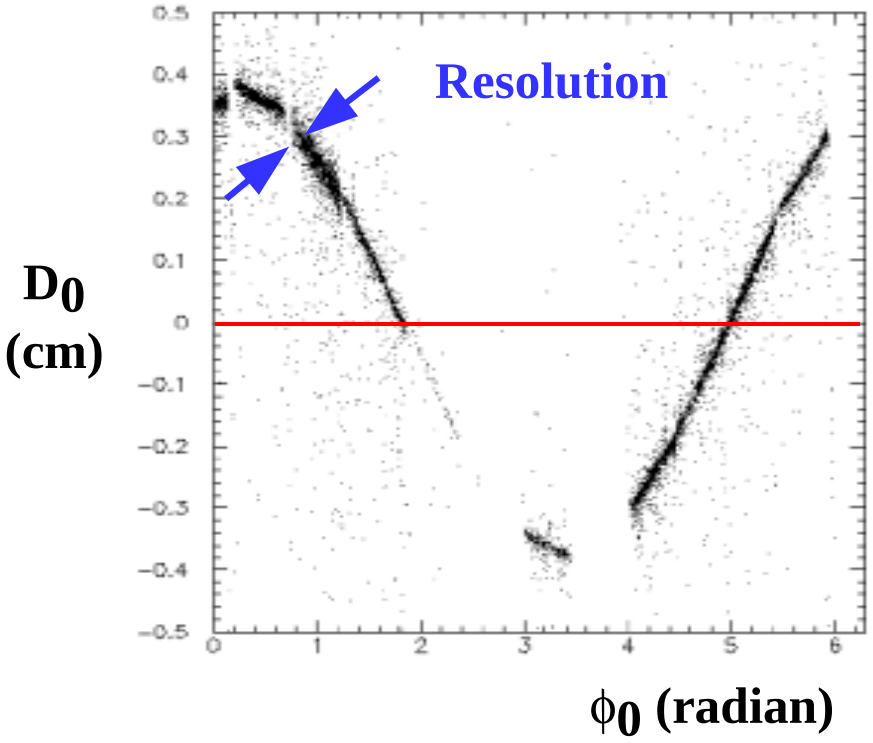


L2: Silicon Vertex Tracker

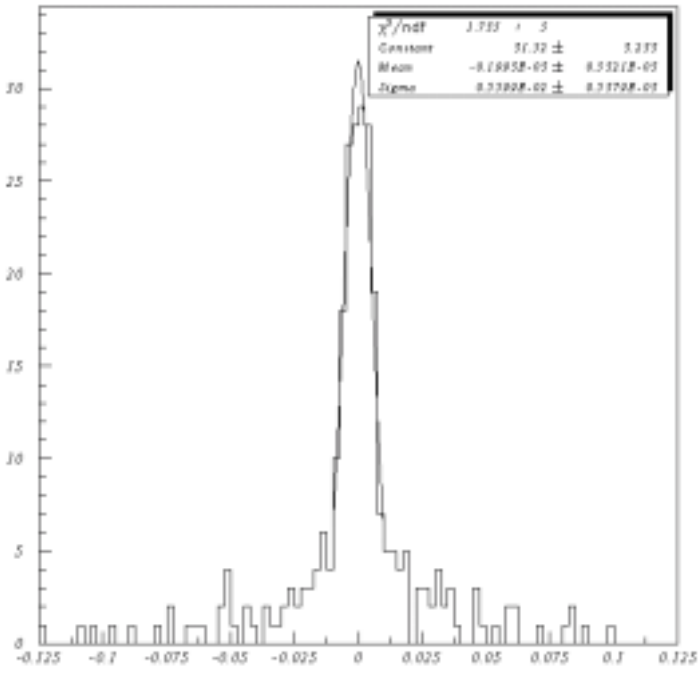
Arnd Meyer
Sep 5, 2001

- Level 2 uses additional/refined information from calorimeter (clusters, isolation), muon systems and tracking (improved matching)
- Trigger algorithms run on 4 (Run IIa) custom Alpha-based VME computers
- Displaced vertex trigger on Level 2
- Hadronic B trigger, e.g. $B_0 \rightarrow \pi^+\pi^-$, $B_S \rightarrow D_S\,\pi$

Level−2 Silicon Trigger measures beam position



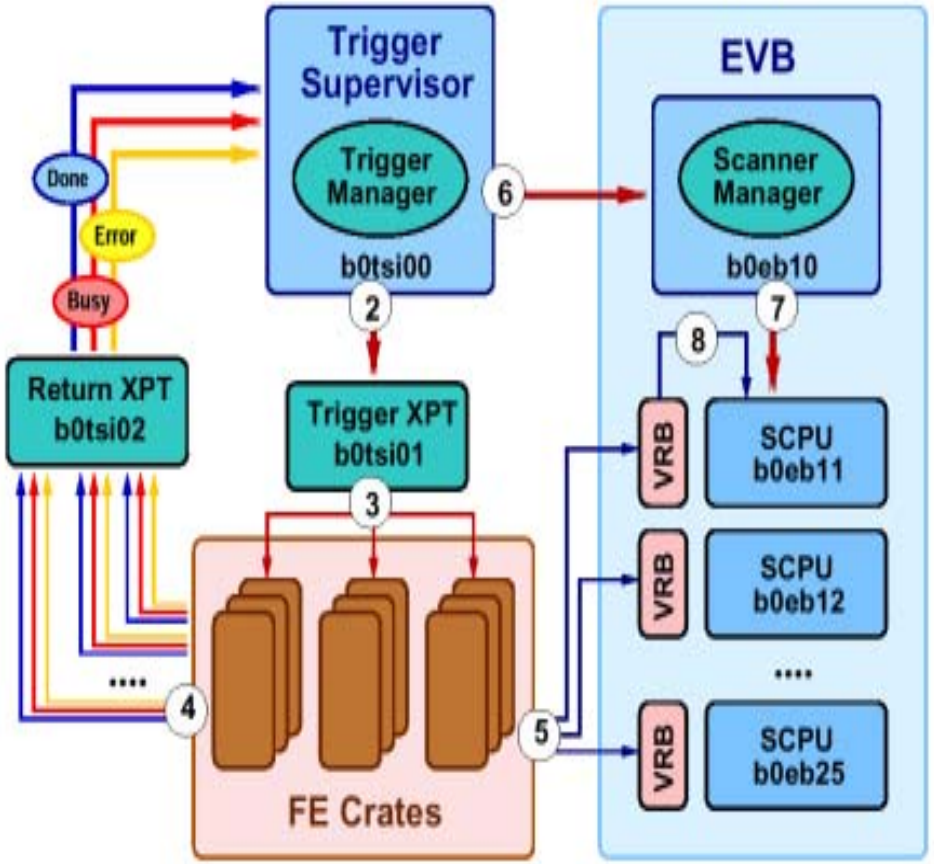
Resolution : 56 μm including the beam spread



Life Cycle of an Event

Arnd Meyer
Sep 5, 2001

- Each bunch crossing, a Level 1 trigger decision is made on the crossing that happened 42 cycles previously
- If accepted, the Level 2 trigger is started (as soon as it is not busy)
- The Level 2 trigger sends back an accept or reject decision when done
- If the Level 2 decision is positive, a message is sent via the Trigger Crosspoints ② to all front-end and trigger VME crates ③
- The processor in each crate reads data from the local modules
  - ④DONE is returned when this is complete
  - If there is some error, a message is sent to the central error handler, and DONE is not set. This will cause data taking to stop
- The crate processor sends the data to the Event Builder ⑤
  - If the VRBs cannot accept more data, BUSY is asserted
  - If BUSY is asserted for more than some timeout, data taking stops
- When DONE is returned by all crates, the Level 2 accept message is forwarded to the Event Builder ⑥



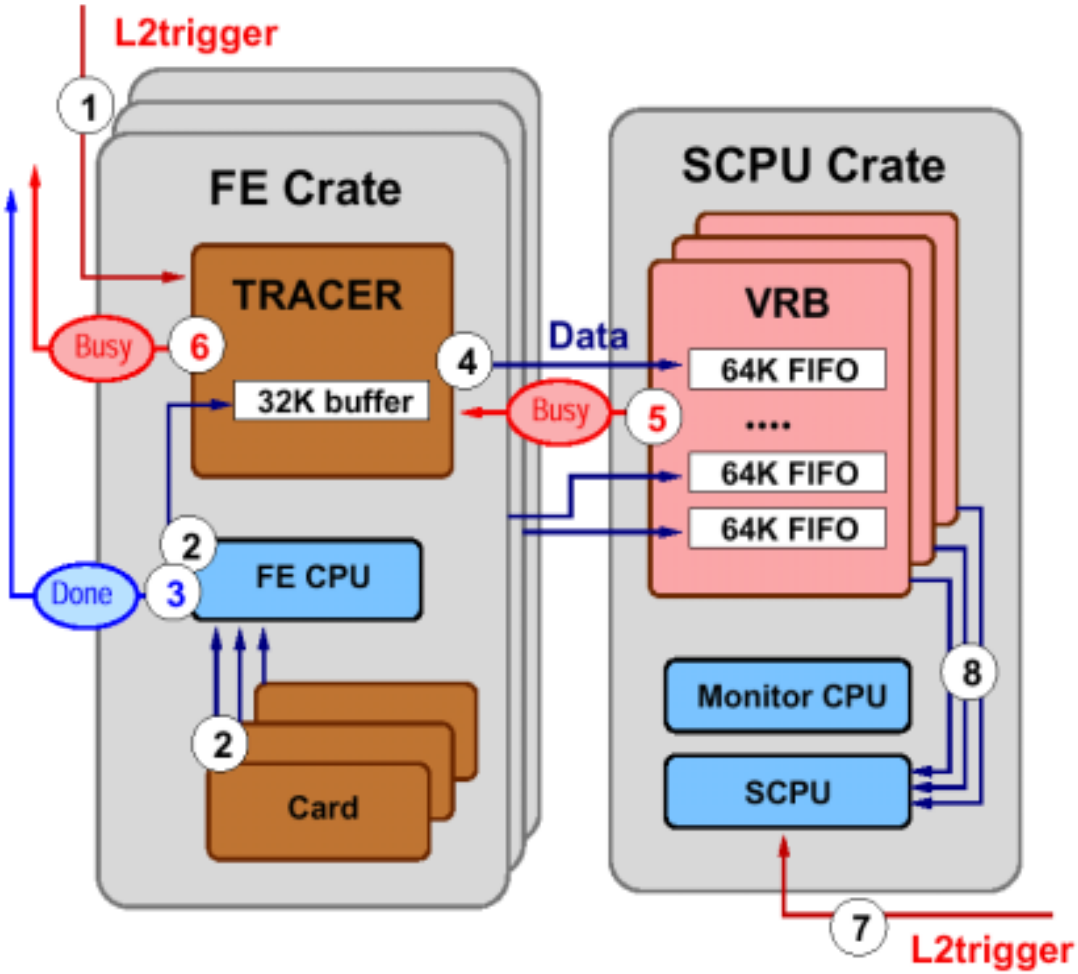
Life Cycle of an Event cont.

Arnd Meyer
Sep 5, 2001

- The Scanner CPUs read data from the VRBs and do integrity checks ®
  - If the data from the VRB is corrupt, the Event Builder will send a message to the error handler and stop reading events. This will cause a BUSY timeout.
- If the data is OK, the data block is sent through the ATM switch to a Level 3 converter node
- The Level 3 converter node concatenates all the fragments into one block and sends it to a Level 3 processor node
- The Level 3 processor node "reformats" the event
  - If corruption is found in the data, a message is sent to the error handler
  - If corrupt, the event is dropped at this point and not processed further
- If the event is not corrupt, the trigger algorithm is run
- If the event passes the trigger, it is passed to an output node
- The output node forwards the data to the CSL
- The CSL writes the data on disk
  - Data will eventually be copied to tape (Sony AIT2) in the computing center
- The CSL may send the event to a "Consumer" process

Life Cycle of an Event cont.

Arnd Meyer
Sep 5, 2001



Trigger System Interface

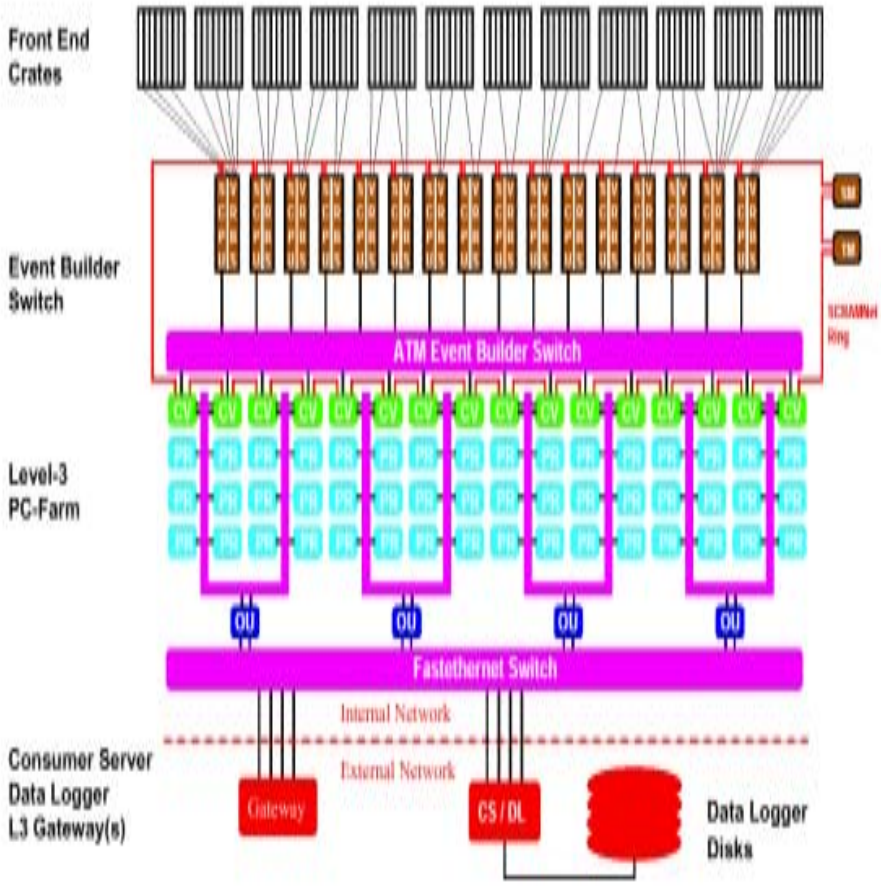
Arnd Meyer
Sep 5, 2001

- Coordinates activity of front-end electronics and trigger from beam crossing through initial phase of readout
- Manage L2 buffer assignment
- Manage live/deadtime
- Trigger Supervisor interacts with L1/L2 global trigger modules to obtain decisions
- Level 1 and 2 accept messages fanned out to front-end systems via Trigger Crosspoints / optical fiber
- Readout status sent back from front-end systems to Trigger Supervisor via Return Crosspoints / copper cables
  - DONE - Readout in progress (false) or done (true)
  - ERROR - Error detected by front-end card before readout - stop data taking
  - BUSY - Buffers in Event builder are full, cannot send data
- Crosspoints allow system to be "partitioned"
  - 8 Trigger Supervisors in total
  - Any single crate may belong to any of 8 partitions (or more)
  - SVX cannot be split however

Event Builder

Arnd Meyer
Sep 5, 2001

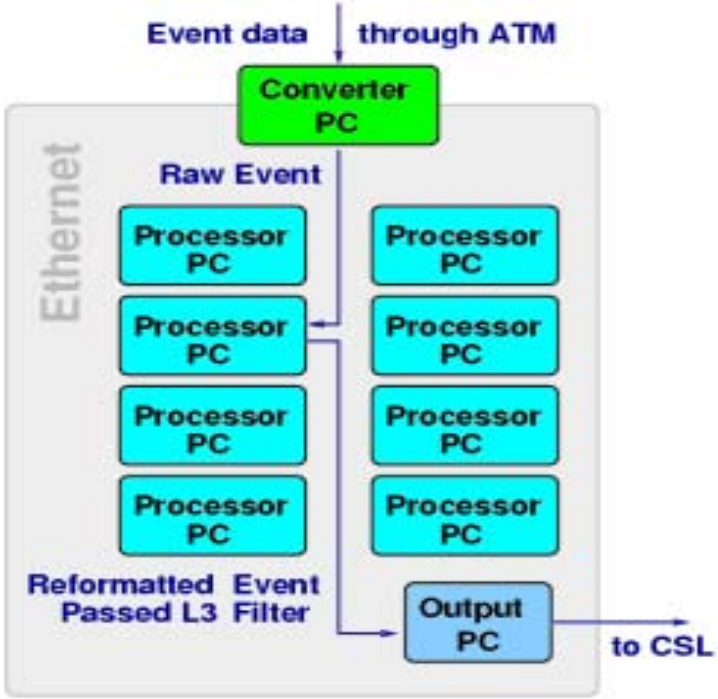
- Collate fragments from each front-end crate into a single block
- Two phases
  - First, data from some number of front-ends are collated in 15 Scanner CPU crates
    - Each of these crates contains one or more VRB (VME Readout Board) modules that contains 10 serial data inputs
    - MVME2603 running VxWorks in each crate collects the data from all local VRBs
  - The 15 fragments are sent through the ATM switch (16in/16out)
  - All 15 fragments wind up in one of 16 Converter nodes in the Level 3 trigger system, where they are concatenated and sent to a Level 3 processor node
- The Scanner Manager sends control messages around to all crates to make sure all fragments are sent to the correct place

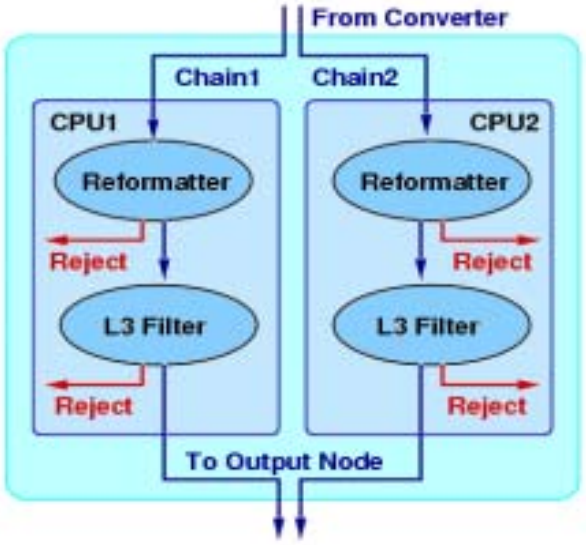


Level 3 Trigger

Arnd Meyer
Sep 5, 2001

- Primary purpose is to apply the final level of filtering to the data
  - Runs programs derived from offline package using the full event data
- "Farm" of dual-processor PCs running Linux, mostly 800MHz Pentium III
  - Currently 16 converter nodes, ~128 processor nodes, 4 output nodes
  - Will be expanded in the fall
- Processor node "reformats" events, sorting by detector component rather than front-end crate
  - Data integrity checks are done at this time
- If the event passed the trigger algorithm based on regular offline (C++) code, it is sent to an output node via Ethernet
- The output node passes it on to the Consumer Server/Logger (CSL)
- Gateway node interfaces private Level 3 ethernet to the public network





"Software" Event Builder

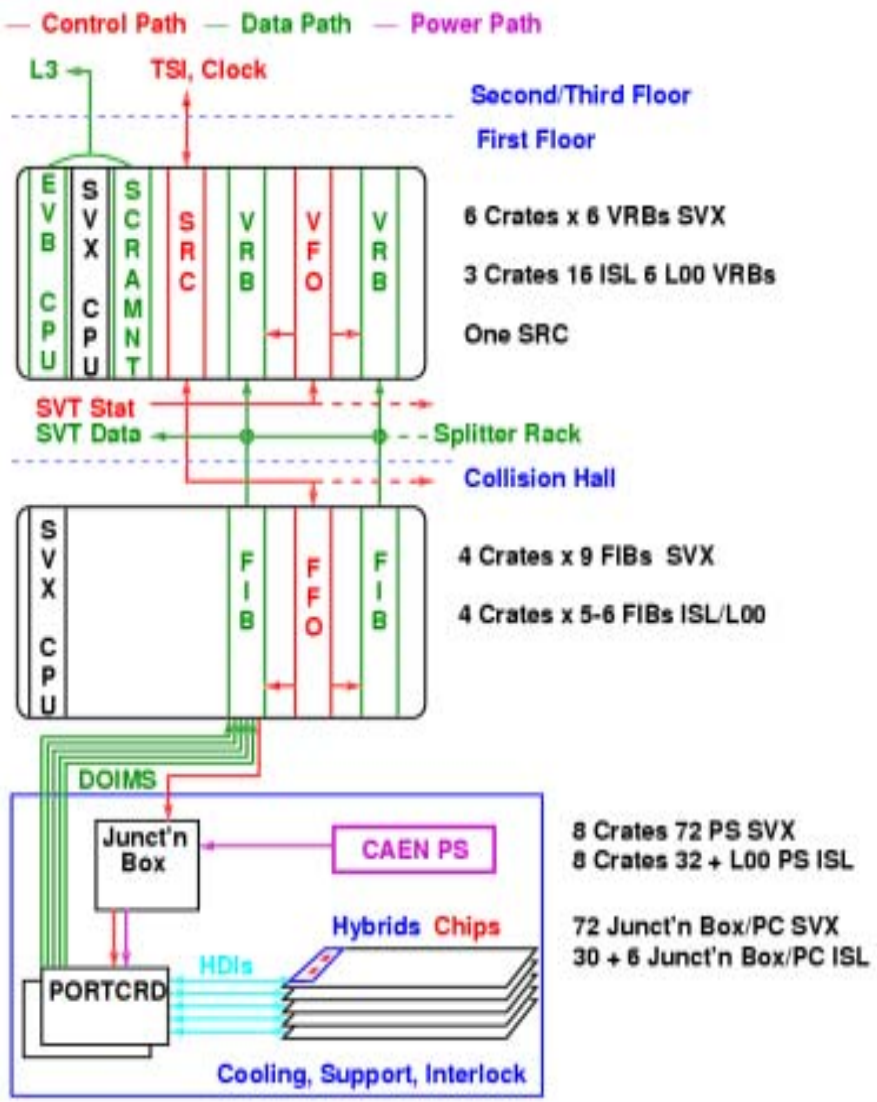
Arnd Meyer
Sep 5, 2001

- Instead of using the complete ATM event builder and Level 3 system, can run with the so-called "Software" event builder
- Data are sent from the crate controllers over ethernet to a single dedicated program
  - Collects all fragments for each events
  - Runs the same "reformat" code as Level 3, but no trigger algorithm
  - Sends assembled event to Consumer Server/Logger
  - At this point it can be treated like any other event
- Obviously much lower data rate (typically <2MB/sec), but used extensively
  - Subsystem testing
  - Calibration runs

Silicon Detector DAQ

Arnd Meyer
Sep 5, 2001

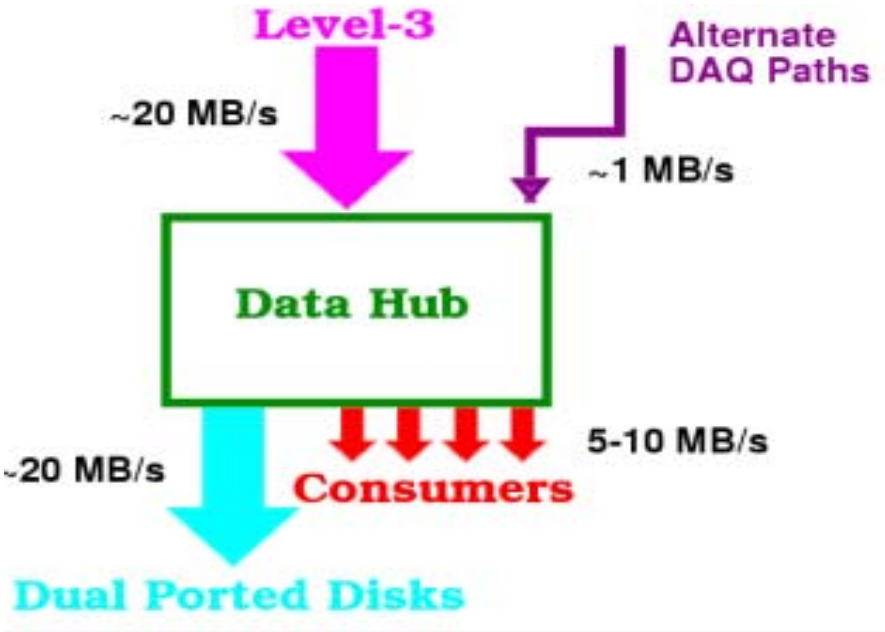
- Some differences compared to rest of detector
- Data flows from detector → FIB crates → SCPU crates in event builder
- Event readout is driven by hardware
  - Crate controller only for configuration and monitoring, does not participate in event acquisition
  - Data is read out on every Level1 accept (not Level2 accept) → can participate in Level2 trigger (SVT)
- Trigger signal fanout done through SRC (Silicon Readout Controller) plus SVX specific fan-out modules
  - No Tracer module

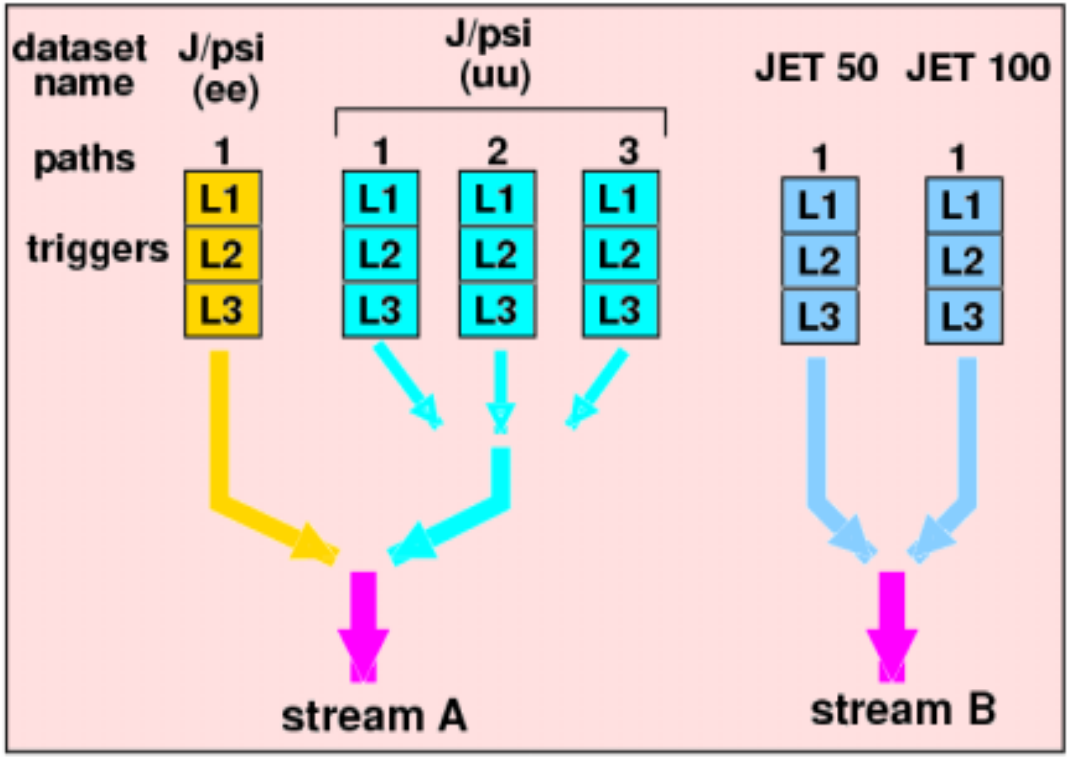


Consumer Server / Logger

Arnd Meyer
Sep 5, 2001

- Set of processes running on a single dedicated SGI machine
- Receives accepted events from Level 3 output nodes via 4 parallel Ethernet ports



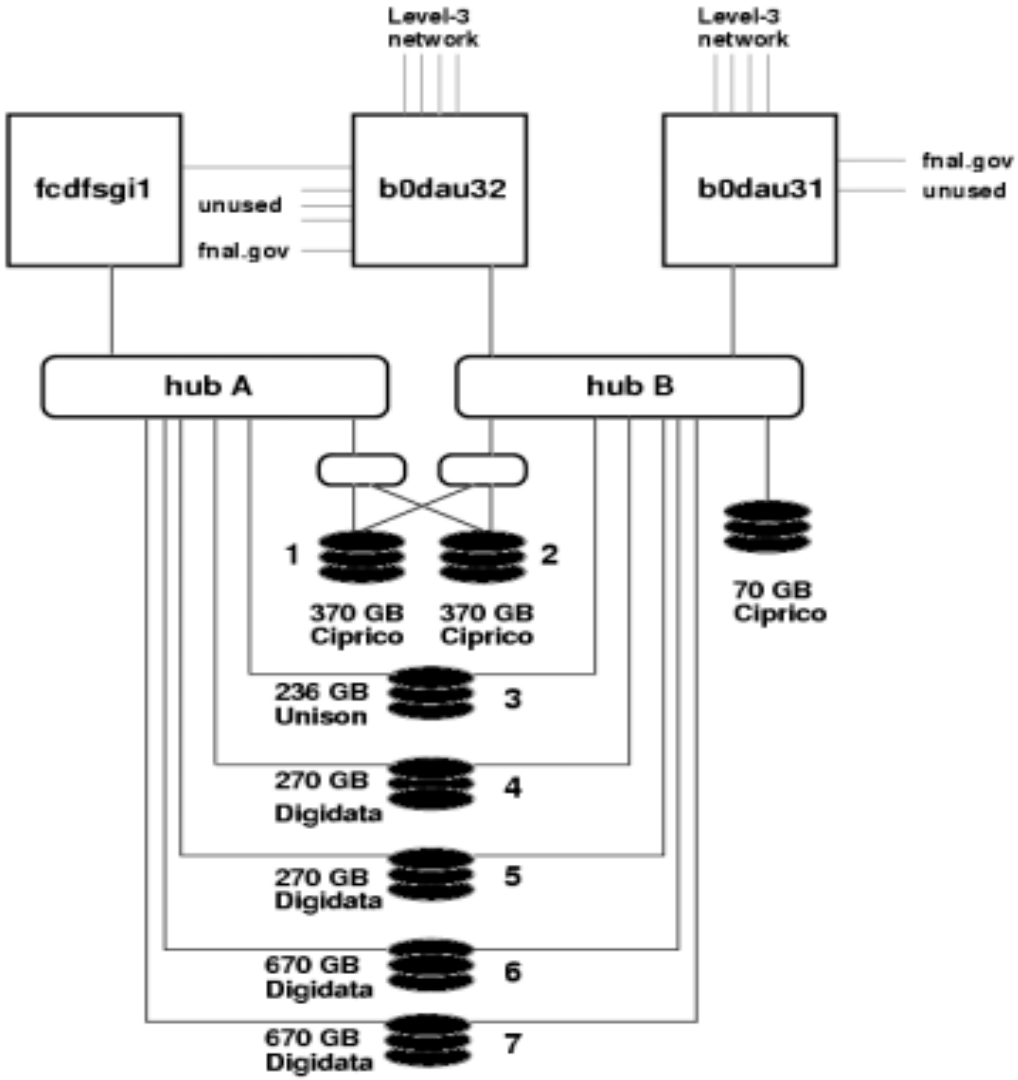


- Writes events to disk
  - Sorted by stream
  - Makes entries in the "Data File Catalog"
- Distributes events to online monitor programs ("Consumers"), see separate talk

CSL cont.

Arnd Meyer
Sep 5, 2001

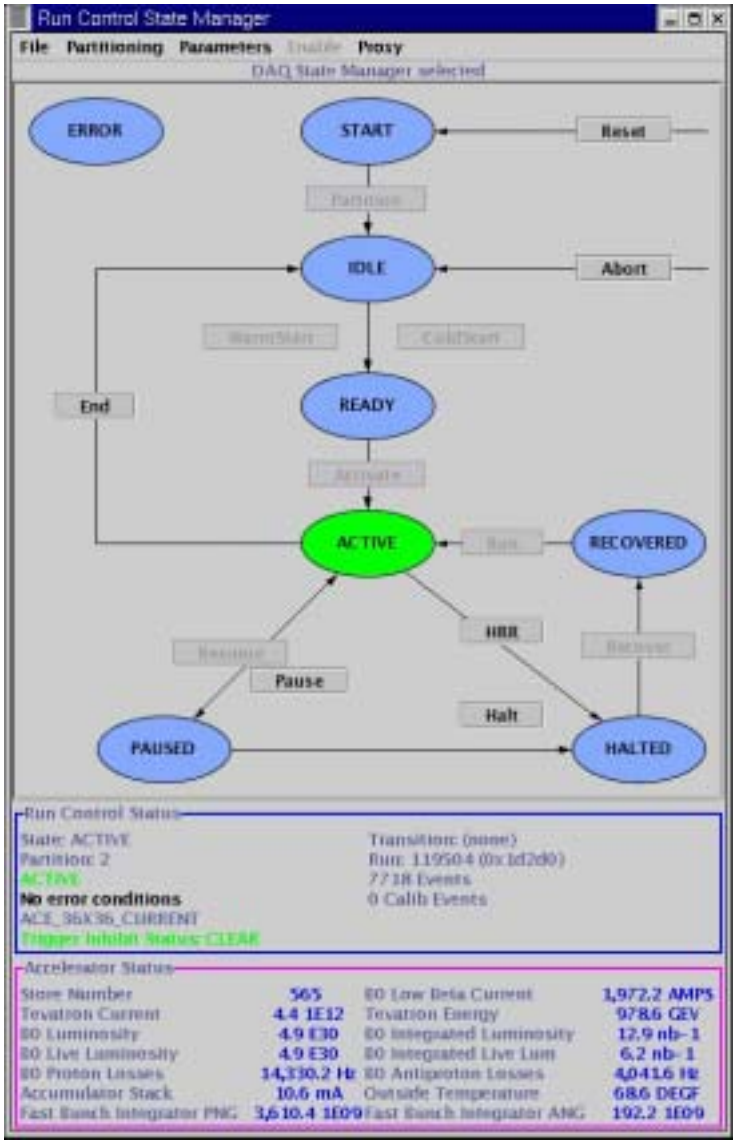
- Disks are dual ported. Tasks in Computing Center read data from 2nd port and write to tape there
- About 2.1TB in use for data logging
- Comfortable for >8h of temporary storage



Run Control / Online Software

Arnd Meyer
Sep 5, 2001

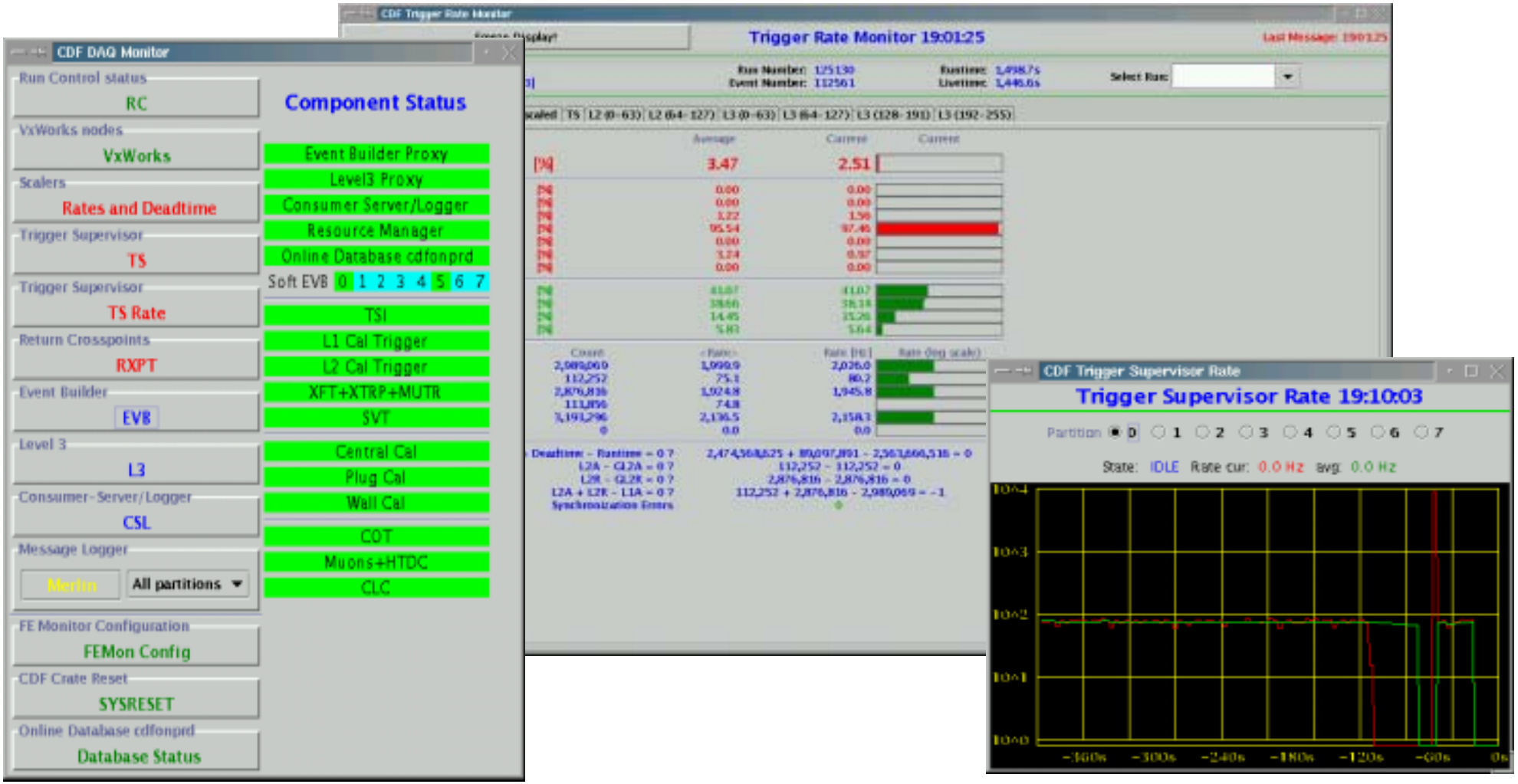
- Run Control is a multi-threaded Java 1.3 application
- About 150 clients written in Java, C (front-end crates), C++ (Level 3)
- Smartsockets from Talarian for communication, using publish/subscribe
  - API on top to synchronize clients written in different languages, simplify use
- Some clients communicate through a Proxy (EVB, Level 3)
- Smartsockets is also used to distribute DAQ monitoring information
  - Java clients in the control room, servlets on the web
- Database API uses JDBC
- Scriptable using JPython
- ~30 control room PCs are mostly Linux (NT for slow control), ~5 file/db servers running IRIX, Linux, Solaris
- Most of the online software can be run anywhere
- Use cvs and Fermilab release tools ups/upd



DAQ Monitoring

Arnd Meyer
Sep 5, 2001

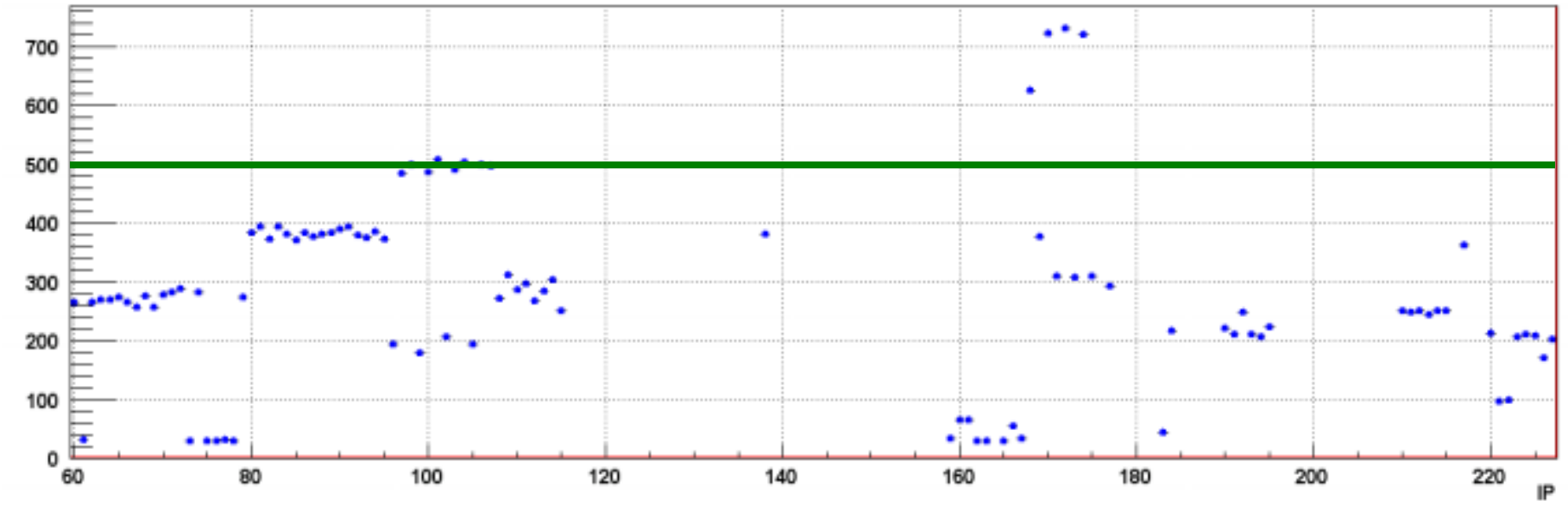
- Variety of monitoring programs written in java, listening to Smartsockets messages and alerting the shift crew
- Central error handler for filtering, analyzing and display of messages, and take appropriate action



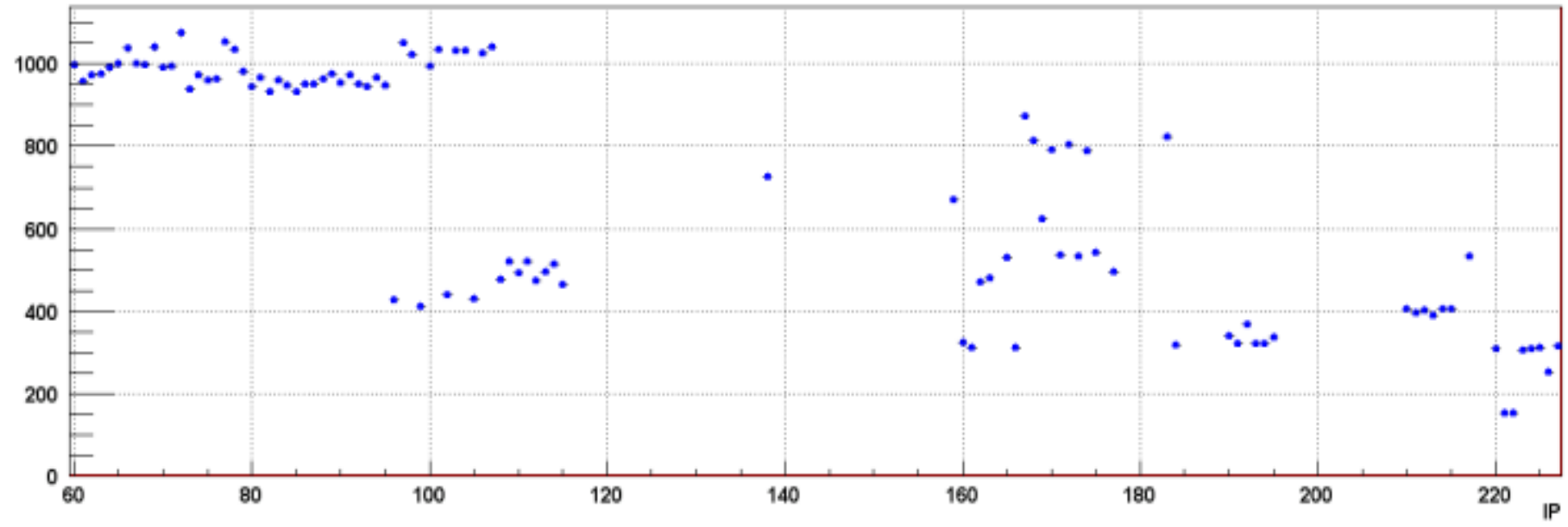
Current Performance – Readout

Arnd Meyer
Sep 5, 2001

Time until buffer freed [μs]



Total readout time [μs]



Current Performance Cont.

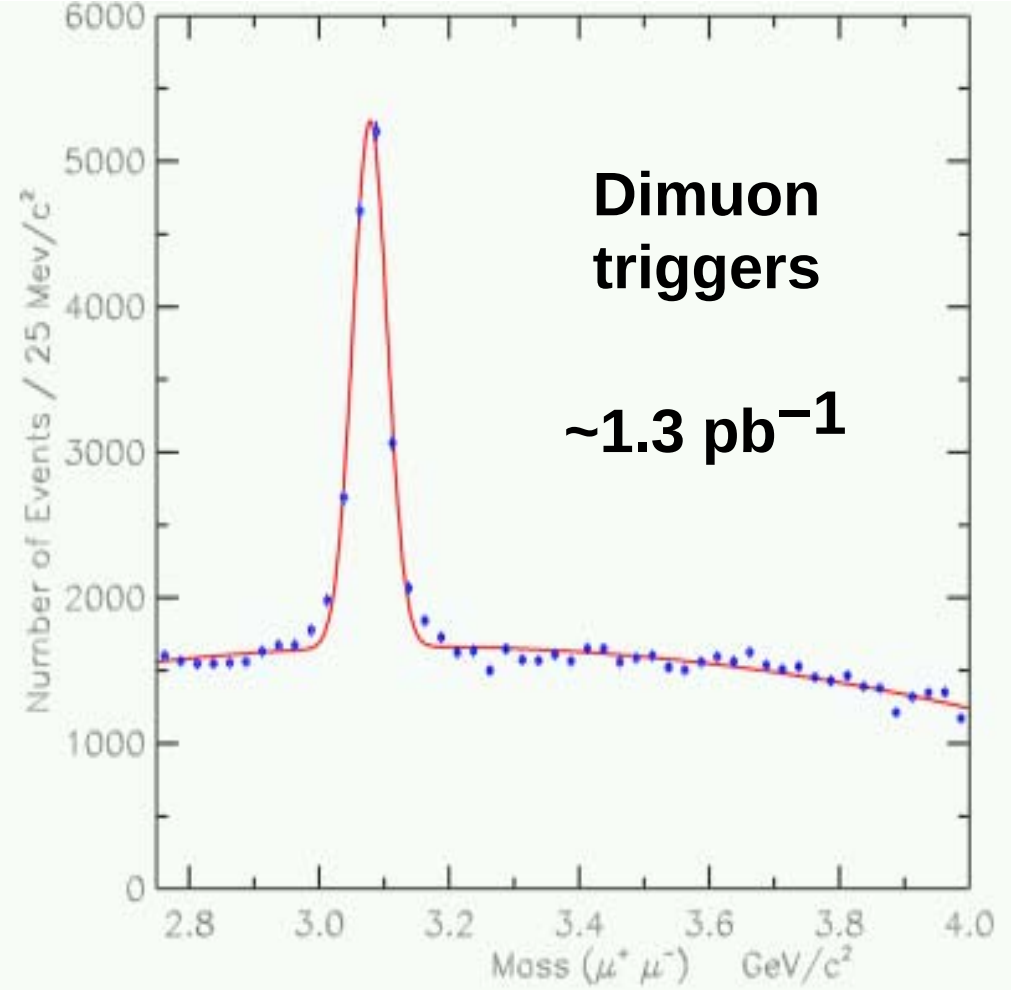
Arnd Meyer
Sep 5, 2001

- Trigger
  - Level 1 fully functional, some pieces missing (parts of muon systems)
  - Level 2 not yet always rejecting, being commissioned: muon trigger missing, calorimeter clustering basically ready, SVT (displaced vertex trigger) waiting
- Event Builder
  - proven up to 400-500Hz
  - May need to upgrade ATM switch for Run IIb, will saturate below 1kHz
- Level3
  - In "reject mode" since July
  - not saturated yet; will be upgraded in fall
- CSL: routinely operated at up to 23MB/sec
- Typically run at 200-300Hz into Level 3, little deadtime (~1%); about 95% live at 360Hz
- Overall data taking efficiency ~50% (still lots of commissioning)
- Peak luminosity > 8 x 10^{30} $cm^{-2}sec^{-1}$
- About 2pb^{-1} of data on tape

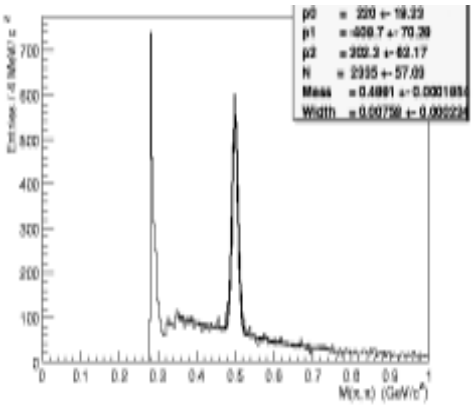
CDF Performance

Arnd Meyer
Sep 5, 2001

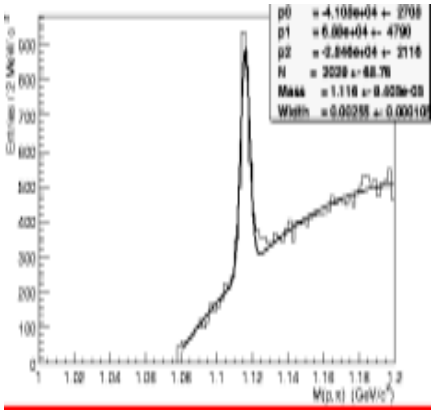
$J/\psi \rightarrow \mu^+\mu^-$



$K_S \rightarrow \pi^+\pi^-$

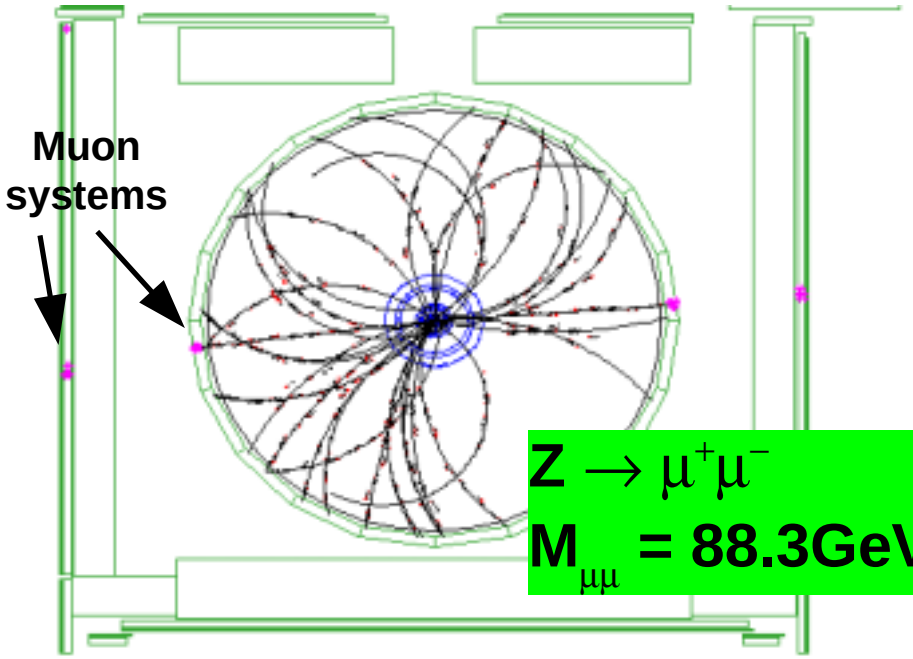


$\Lambda \rightarrow \pi^-p$

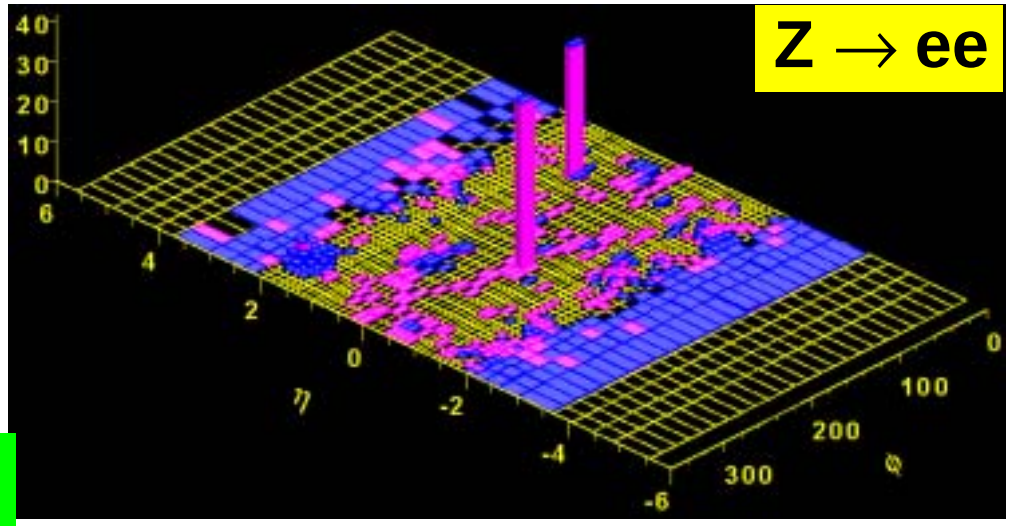


Z Candidates

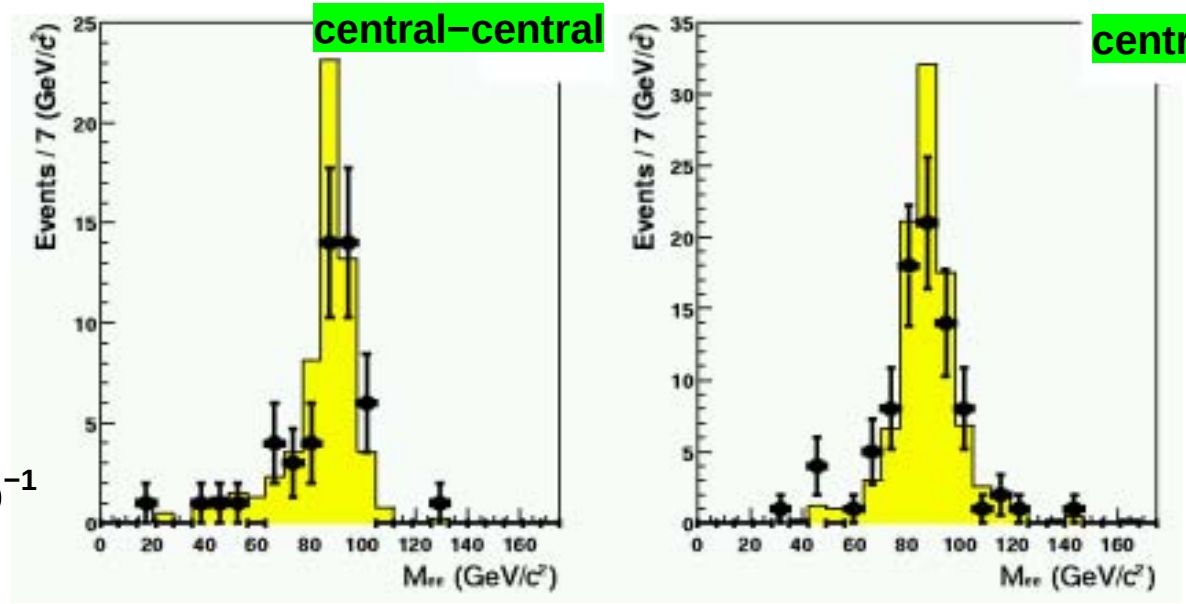
Arnd Meyer
Sep 5, 2001



$Z \rightarrow \mu^+\mu^-$
$M_{\mu\mu}$ = 88.3GeV



~1.8pb^{-1}



Summary

Arnd Meyer
Sep 5, 2001

- The CDF DAQ had a good start into run II
- Most of the design specifications have already been reached or surpassed
- Taking collision data since April, now close to/at "physics quality data"
- Upcoming ~6 week shutdown in October/November, uninterrupted running after that
- Expect ~200pb^{-1} by summer 2002
- Detector fully operational (except 1 (out of 7) silicon layer has partial cooling problem)
- Level 1+3 triggers in good shape, Level 2 being commissioned